Google Scholar BETA

Web Images Video News Maps **more »**

("edit distance" OR "editing distance" OR "leve Search

Advanced Scholar Search
Scholar Preferences
Scholar Help

The **"AND"** operator is unnecessary -- we include all search terms by default. [details]

Scholar **All articles** - **Recent articles** Results **1** - **10** of about **749** for **("edit distance" OR "editing di**

All Results
G Navarro
J Schmidt
E Myers
G Landau
S Uliel

[CITATION] Incremental String Comparison - all 11 versions »
GM Landau, EW Myers, JP Schmidt - SIAM J. Comput., 1998
... of D as an (m + 1) × (n + 1) **grid** of points (i ... O(km) algorithm to determine the k-thresholded **edit distance** of A ... because if we were to extend the **matrix** D with ...
Cited by 44 - Related Articles - Web Search - BL Direct

Naturally occurring circular permutations in proteins - all 4 versions »
S Uliel, A Fliess, R Unger - Protein Engineering Design and Selection, 2001 - Oxford Univ Press
... right side of the last **row**, ie of **columns** N + 1 ... cell of the last **row** of the **edit distance matrix** was smaller ... of the average value of all cells in the last **row**. ...
Cited by 20 - Related Articles - Web Search - BL Direct

[CITATION] All highest scoring paths in weighted **grid** graphs and their application to finding all approximate ... - all 7 versions »
JP Schmidt - SIAM J. Comput, 1998
... 94] using a general penalty **matrix**, and, among ... tandem repeats under the **Levenshtein distance**, in which ... and report repeated patterns whose **edit distance** is less ...
Cited by 53 - Related Articles - Web Search - BL Direct

A simple algorithm for detecting circular permutations in proteins - all 8 versions »
S Uliel, A Fliess, A Amir, R Unger - Bioinformatics, 1999 - Oxford Univ Press
... and the approximate algorithm (as read from the last **row** of the doubled **matrix**) for two ... even the exact algorithm (ie calculating the **edit distance** for each ...
Cited by 24 - Related Articles - Web Search

Approximate Matching of Run-Length Compressed Strings - all 10 versions »
V Mäkinen, E Ukkonen, G Navarro - Algorithmica, 2003 - Springer
... For the **Levenshtein distance** (denoted by D L (A, B)) [18 ... In general, the **edit distance** D(A, B) with arbitrary δ ... The **matrix** (d ij) can be evaluated **row**-by-**row** ...
Cited by 15 - Related Articles - Web Search - BL Direct

A guided tour to approximate string matching - all 18 versions »
G Navarro - ACM Computing Surveys (CSUR), 2001 - portal.acm.org
... search and retrieval—Search process General Terms: Algorithms Additional Key Words and Phrases: **Edit distance**, **Levenshtein distance**, online string matching ...
Cited by 358 - Related Articles - Web Search - BL Direct

Edit distance string search
ET Bax, ID Swett - 2004 - freepatentsonline.com
... 5a) Get the **edit distance**, which is the value in the **grid** cell on the **row** corresponding to ... If the **edit distance** is less than the **threshold**, then report the ...
Cached - Web Search

Adaptively weighted, partitioned context **edit distance** string matching
AC Richardson, CM Davis, DP Miranker - 2002 - freepatentsonline.com
... [0044] The **edit distance** matrices contain values representing the edit distances ... Put another way, each **matrix** cell has a corresponding **row** and **column**. ...
Cached - Web Search

APPROXIMATE STRING MATCHING BY COMBINING AUTOMATON APPROACH AND BINARY NEURAL NETWORKS Toma˘s Beran ...
M Skrbek, T Macek - actapress.com
... a single error (consid- ering the Damerau-**Levenshtein distance**). ... algorithms for Damerau-Levenshtein **edit distance** k = 1 ... combining with the index **matrix** method. ...
Web Search

[PS] INCREMENTAL STRING COMPARISON - all 4 versions »
M GAD, WM EUGENE, P JEANETTE - cs.haifa.ac.il
... The reader may wish to also verify that the **edit distance** problem is ... that it is easy to extend D to a **matrix** D 0 ... A versus Bb by computing an additional **column**. ...
Related Articles - View as HTML - Web Search

Gooooooooogle ▶
Result Page: 1 2 3 4 5 6 7 8 9 10 **Next**

```
Set     Items   Description
S1      87348   (WORD??? OR CHARACTER? OR TERM? ? OR TEXT? ? OR ALPHABET? -
             OR SPELLING? OR LETTER?)(3N)(STRING? OR SEQUEN? OR PATTERN? OR
              COMBINATION?)
S2       5678   S1(7N)(MONITOR? OR EXAMIN? OR DETECT? OR UNCOVER? OR REVEA-
             L? OR ASSESS? OR EVALUAT? OR INSPECT?)
S3      13414   S1(7N)(DETERMIN? OR COMPAR? OR DISCERN? OR ASCERTAIN? OR A-
             NALY? OR IDENT? OR CHECK? OR VERIF? OR JUDG???)
S4       8226   S1(5N)(SET? ? OR GROUP? OR CLUSTER? OR BUNCH? OR COLLECTIO-
             N? OR AGGREGAT? OR SAMPL? OR DICTIONAR?)
S5     380681   (MAXIMUM? OR MAX OR TOP OR HIGH? OR GREAT? OR PEAK??? OR P-
             ENULTIMAT? OR FULL??? OR MOST?? OR LAST?? OR THRESHOLD? OR TH-
             RESHHOLD? OR BRINK)(5N)(NUMBER? OR AMOUNT? OR COUNT? OR SUM OR
              TOTAL? OR TALLY? ? OR QUANTIT? OR MANY OR KNOWN OR SIZE? ?)
S6      10341   S5(7N)(EDIT??? OR LEVENSHTEIN? OR HAMMING? OR PARE? ? OR P-
             ARING OR INSERT? OR DELET? OR TRANSFORM? OR CHANG???)
S7      97302   (STOP??? OR STOPPING OR CEASE?? OR CEASING OR END??? OR HA-
             LT??? OR CESSAT?)(3N)(COMPUT? OR DETERMIN? OR TOTAL? OR TABUL-
             AT? OR CALCULAT? OR PROCESS? OR FIGUR? OR FORMULAT?)
S8       2755   S2:S3 AND S4
S9        176   S8 AND S5
S10         5   S9 AND S6
S11        28   S9 AND (MATRIX? OR GRID? OR TABLE? OR MATRIC?)
S12         0   S11 AND (EDIT??? OR LEVENSHTEIN?)
S13        98   S8 AND (EDIT??? OR LEVENSHTEIN?)
S14        11   S13 AND (MATRIX? OR GRID? OR TABLE? OR MATRIC?)
S15        11   S14 NOT S10
S16        29   S S1:S3 AND S5 AND (EDIT??? OR LEVENSHTEIN?)
S17         0   S16 AND LEVENSHTEIN?
S18         1   S16 AND AC=US/PR AND AY=(2004:2007)/PR
S19         5   S16 AND AC=US AND AY=2004:2007
S20         1   S16 AND AC=US AND AY=(2004:2007)/PR
S21         9   S16 AND PY=2004:2007
S22         9   S18:S21
S23        20   S16 NOT S22
S24        20   S23 NOT (S10 OR S14:S15)
S25         6   LEVENSHTEIN?
S26        21   EDIT?()DISTAN? AND (S1:S3 OR S8)
S27        20   S26 NOT (S10 OR S14:S15 OR S22:S25)
File 350:Derwent WPIX 1963-2007/UD=200762
         (c) 2007 The Thomson Corporation
File 347:JAPIO Dec 1976-2007/Jun(Updated 070926)
         (c) 2007 JPO & JAPIO
```

27/69,K/6 (Item 6 from file: 350)
DIALOG(R)File 350:Derwent WPIX

0014649586
WPI ACC NO: 2004-831605/200482
XRPX Acc No: N2004-657016
Edit distance **string search method for e.g. to locate medical history of patient, involves setting next computation text as next to present computation text that does not share forbidden prefix, if all columns are not computed**
Patent Assignee: BAX E T (BAXE-I); SWETT I D (SWET-I)
Inventor: BAX E T; SWETT I D
Patent Family (1 patents, 1 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
US 20040220920	A1	20041104	US 2003449007	P	20030224	200482 B
			US 2004775576	A	20040209	

Priority Applications (no., kind, date): US 2003449007 P 20030224; US 2004775576 A 20040209

Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
US 20040220920	A1	EN	11	0	Related to Provisional US 2003449007

Alerting Abstract US A1

NOVELTY - The column-by-column grid computation of the **edit distance** is performed between search **string** and computation **text** . If all columns are not computed, the prefix of the text corresponding to the columns up to and including the column with minimum threshold is made as forbidden prefix so that next computation text is set as next to present computation text that does not share the forbidden prefix.

USE - For searching **edit distance** string e.g. to locate the medical history of a patient among large set of histories, to detect duplicate benefit payments by government agencies, and to aggregate data about a customer with multiple accounts at a financial institution.

ADVANTAGE - Uses column bounding and prefix column sharing to reduce computation, while increasing computation speed.

Title Terms/Index Terms/Additional Words: EDIT; DISTANCE; STRING; SEARCH; METHOD; LOCATE; MEDICAL; HISTORY; PATIENT; SET; COMPUTATION; TEXT; PRESENT; SHARE; FORBID; PREFIX; COLUMN

Class Codes
International Classification (Main): G06F-007/00
US Classification, Issued: 707003000

File Segment: EPI;
DWPI Class: T01
Manual Codes (EPI/S-X): T01-J05B3; T01-J06A1

Edit distance **string search method for e.g. to locate medical history of patient, involves setting next...**

Original Titles:
Edit distance string search

Alerting Abstract ...NOVELTY - The column-by-column grid computation of the **edit distance** is performed between search **string** and computation

text . If all columns are not computed, the prefix of the text corresponding to the columns...
USE - For searching **edit distance** string e.g. to locate the medical history of a patient among large set of...

Original Publication Data by Authority

Original Abstracts:
A process determines for a search **string** which, if any, of the **strings** in a **text list** have **edit distance** from the **search string** less than a threshold. The process uses dynamic programming on a grid with search **string characters** corresponding to **rows and text characters** corresponding to **columns** . **For** each **text string** , computation proceeds **by columns** . If successive **text strings** share a **prefix** , **then** the columns corresponding to the prefix are re-used. If the minimum value in a...

...is at least the threshold, then the prefix corresponding to that and previous columns causes **edit distance** to be **at least** the threshold. So the computation for the present text is abandoned, and computations for any...
Claims:
...claimed is: 1. A method for identifying among a list of texts those that have **edit distance** from a search **string** that is less **than a** threshold, said method comprising: beginning with a first text from a list of texts as a computation **text** ; providing **a** search **string** and a **threshold** value; performing a column-by-column grid computation of **edit distance** between the search **string** and the computation **text** , stopping early if a column **minimum value** is at least **the** threshold; if the **edit distance** is computed and is below the threshold, then reporting that the computation text is close **to the** search string in **edit distance** ; if all columns are computed, then setting the next computation **text** to the **next text in the** list after the present computation text; if not all columns are computed, then setting the...

27/69,K/8 (Item 8 from file: 350)
DIALOG(R)File 350:Derwent WPIX

0013668333 - Drawing available
WPI ACC NO: 2003-764701/200372
XRPX Acc No: N2003-612426
Word/phrase spelling correction method involves accepting replacement word, if its actual distance from given word for spelling correction is less than maximal edit distance
Patent Assignee: INT BUSINESS MACHINES CORP (IBMC)
Inventor: BIRMAN A; GAIL H R; HANTLER S L; LEEMAN G B; MILCH D
Patent Family (1 patents, 1 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
US 6616704	B1	20030909	US 2000665897	A	20000920	200372 B

Priority Applications (no., kind, date): US 2000665897 A 20000920

Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
US 6616704	B1	EN	5	1	

Alerting Abstract US B1

NOVELTY - A vector having length equal to number of letters in alphabet for construction of given word to be corrected, is computed, and the replacement word letters are iterated by defining the number of letters in replacement/given words not appearing in given/replacement words. The replacement word is accepted, if its actual distance from the given word is less than maximal **edit distance** .

DESCRIPTION - An INDEPENDENT CLAIM is also included for computer-readable medium storing codes for correcting spelling of word or phrase in document.

USE - For correcting spelling of words or phrases in documents.

ADVANTAGE - The spelling of word or phrase in a document, is corrected quickly and exactly.

DESCRIPTION OF DRAWINGS - The figure shows the flowchart explaining the steps involved for correcting the spelling of word or phrase in document.

Title Terms/Index Terms/Additional Words: WORD; PHRASE; SPELLING; CORRECT; METHOD; ACCEPT; REPLACE; ACTUAL; DISTANCE; LESS; MAXIMUM; EDIT

Class Codes
International Classification (Main): G06F-017/21
US Classification, Issued: 715533000, 707531000, 707532000

File Segment: EPI;
DWPI Class: T01
Manual Codes (EPI/S-X): T01-J11A1; T01-S03

...word, if its actual distance from given word for spelling correction is less than maximal edit distance
...word is accepted, if its actual distance from the given word is less than maximal **edit distance** .

Original Publication Data by Authority

Original Abstracts:
...very fast approximate method for eliminating most candidate words from consideration (without computing the exact **edit distance** between the

given word whose spelling is to be corrected and any candidate word), followed by a "slow method" which computes the exact **edit distance** between the **word whose** spelling is to be corrected and each of the few remaining candidate **words** . The **combination** results **in** a **method** that is almost as fast as the fast approximate method and as exact as the...

Claims:

...applying an approximate method for eliminating some candidate words from consideration, without computing an exact **edit distance** between a given word whose spelling is **to be** corrected and any candidate word; followed by applying an exact method which computes an exact **edit distance** between the word whose spelling is to be corrected **and each** of the remaining candidate words, wherein G is a given word whose spelling is to ...

...entry is zero, incrementing Cval by one; determining if Cval is larger than a maximal **edit distance** Delta, rejecting C from consideration; and continuing letter by letter until either C has been...

...Gval is larger than Delta, then rejecting C from consideration, but otherwise, accepting C, provided **its actual** distance from G is not more than Delta.

27/69,K/11 (Item 11 from file: 350)
DIALOG(R)File 350:Derwent WPIX

0010379920 - Drawing available
WPI ACC NO: 2000-137445/200012
XRPX Acc No: N2000-102749
Search algorithm in search system used for retrieving text information in world wide web
Patent Assignee: FAST SEARCH & TRANSFER ASA (FAST-N); OVERTURE SERVICES INC (OVER-N)
Inventor: RISVIK K M
Patent Family (22 patents, 85 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update	
WO 2000003315	A2	20000120	WO 1999NO233	A	19990709	200012	B
NO 199803175	A	20000111	NO 19983175	A	19980710	200014	E
NO 199903413	A	20000111	NO 19993413	A	19990709	200014	E
AU 199949370	A	20000201	AU 199949370	A	19990709	200028	E
EP 1095326	A1	20010502	EP 1999933296	A	19990709	200125	E
			WO 1999NO233	A	19990709		
BR 199912015	A	20010410	BR 199912015	A	19990709	200128	E
			WO 1999NO233	A	19990709		
CZ 200100064	A3	20010711	WO 1999NO233	A	19990709	200147	E
			CZ 200164	A	19990709		
NO 311657	B1	20011227	NO 19993413	A	19990709	200206	E
CN 1317114	A	20011010	CN 1999810507	A	19990709	200207	E
KR 2001071841	A	20010731	KR 2001700436	A	20010110	200208	E
EP 1095326	B1	20020130	EP 1999933296	A	19990709	200209	E
			WO 1999NO233	A	19990709		
DE 69900854	E	20020314	DE 69900854	A	19990709	200226	E
			EP 1999933296	A	19990709		
			WO 1999NO233	A	19990709		
US 6377945	B1	20020423	WO 1999NO233	A	19990709	200232	E
			US 2000486726	A	20000309		
JP 2002520712	W	20020709	WO 1999NO233	A	19990709	200259	E
			JP 2000559494	A	19990709		
HU 200201630	B	20020828	WO 1999NO233	A	19990709	200264	E
			HU 20021630	A	19990709		
ES 2173752	T3	20021016	EP 1999933296	A	19990709	200279	E
KR 414236	B	20040107	WO 1999NO233	A	19990709	200427	E
			KR 2001700436	A	20010110		
AU 772525	B2	20040429	AU 199949370	A	19990709	200457	E
JP 3581652	B2	20041027	WO 1999NO233	A	19990709	200470	E
			JP 2000559494	A	19990709		
AU 2004203480	A1	20040826	AU 2004203480	A	20040729	200476	NCE
IN 200100144	P4	20050520	WO 1999NO233	A	19990709	200572	E
			IN 2001CN144	A	20011031		
CA 2337079	C	20060704	CA 2337079	A	19990709	200645	E
			WO 1999NO233	A	19990709		

Priority Applications (no., kind, date): AU 2004203480 A 20040729; NO 19983175 A 19980710

Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
WO 2000003315	A2	EN	31	7	

National Designated States,Original: AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GD GE GH HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZA ZW

Regional Designated States,Original: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SL SZ UG ZW

AU 199949370	A	EN		Based on OPI patent WO 2000003315
EP 1095326	A1	EN		PCT Application WO 1999NO233
				Based on OPI patent WO 2000003315

Regional Designated States,Original: AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE

BR 199912015	A	PT		PCT Application WO 1999NO233
				Based on OPI patent WO 2000003315
CZ 200100064	A3	CS		PCT Application WO 1999NO233
				Based on OPI patent WO 2000003315
NO 311657	B1	NO		Previously issued patent NO 9903413
EP 1095326	B1	EN		PCT Application WO 1999NO233
				Based on OPI patent WO 2000003315

Regional Designated States,Original: AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE

DE 69900854	E	DE		Application EP 1999933296
				PCT Application WO 1999NO233
				Based on OPI patent EP 1095326
				Based on OPI patent WO 2000003315
US 6377945	B1	EN		PCT Application WO 1999NO233
				Based on OPI patent WO 2000003315
JP 2002520712	W	JA	74	PCT Application WO 1999NO233
				Based on OPI patent WO 2000003315
HU 200201630	B	HU		PCT Application WO 1999NO233
				Based on OPI patent WO 2000003315
ES 2173752	T3	ES		Application EP 1999933296
				Based on OPI patent EP 1095326
KR 414236	B	KO		PCT Application WO 1999NO233
				Previously issued patent KR 2001071841
				Based on OPI patent WO 2000003315
AU 772525	B2	EN		Previously issued patent AU 9949370
				Based on OPI patent WO 2000003315
JP 3581652	B2	JA	28	PCT Application WO 1999NO233
				Previously issued patent JP 2002520712
				Based on OPI patent WO 2000003315
AU 2004203480	A1	EN		Division of patent AU 772525
IN 200100144	P4	EN		PCT Application WO 1999NO233
CA 2337079	C	EN		PCT Application WO 1999NO233
				Based on OPI patent WO 2000003315

Alerting Abstract WO A2

NOVELTY - Two algorithms determine degree of matching between words in a suffix tree ST(T) of the text T and query Q, and between the **sequences** of **words** . The algorithms search the data structure with queries Q in the form of **words** , **sequences** of **words** or **sequences** of symbols or combinations, so that information R is retrieved on the basis of query with specified degree of matching between the **words** and **word sequences** .

DESCRIPTION - The information retrieval takes place with a given or varying degree of matching between query Q and retrieved information R. A data structure stores text T, metric M which measures the degree of matching between query and retrieved information. A search algorithm executes full text search, based on keywords. The tree structure in the form of suffix tree ST(T) which stores the suffixes of **words** , **word sequences** and symbol **sequences** in a **text** T and the metric having combination of **edit distance** metric for matching between words or

symbols in the text and the query, and between **sequences** S of **words** and symbols and the query sequence P. One of the edit metric has weighting cost functions for edit operations, for transforming sequence S to sequence P. An INDEPENDENT CLAIM is also included for text information retrieval method.

USE - In search system for text information retrieval. Also used to search documents in world wide web (WWW).

ADVANTAGE - Fast and efficient search and retrieval of information in large volume of data is enabled. Can be applied to searching and retrieving information stored in the form of digitalized images and graphic symbols.

DESCRIPTION OF DRAWINGS - The figure shows schematic structure of search engine with search system.

Title Terms/Index Terms/Additional Words: SEARCH; ALGORITHM; SYSTEM; RETRIEVAL; TEXT; INFORMATION; WORLD; WIDE; WEB

Class Codes

International Classification (Main): G06F-001/00, G06F-017/30, G06F (Additional/Secondary): H04L

International Classification (+ Attributes)

IPC + Level Value	Position	Status	Version
G06F-0017/30 A	I	F	20060101
G06F-0017/30 A	I	R	20060101
G06F-0017/30 C	I	R	20060101

US Classification, Issued: 707005000, 707006000, 707007000, 707010000, 707513000, 707003000

File Segment: EPI;
DWPI Class: T01
Manual Codes (EPI/S-X): T01-J05B2B; T01-J05B3; T01-S01C

Alerting Abstract ...a suffix tree ST(T) of the text T and query Q, and between the **sequences** of **words** . The algorithms search the data structure with queries Q in the form of **words** , **sequences** of **words** or **sequences** of symbols or combinations, so that information R is retrieved on the basis of query with specified degree of matching between the **words** and **word sequences** . ...tree structure in the form of suffix tree ST(T) which stores the suffixes of **words** , **word sequences** and symbol **sequences** in a **text** T and the metric having combination of **edit distance** metric for matching between words or symbols in the text and the query, and between **sequences** S of **words** and symbols and the query sequence P. One of the edit metric has weighting cost...

Original Publication Data by Authority

Original Abstracts:

...a data structure for storing a text T, a combined metric M which includes an **edit distance** metric for **approximate degree** of matching between words and/or symbols **or sequences** thereof in the **text** T and **words and** /or symbols **in** a **sequence** P, weighting cost functions for edit operations which transform a **sequence** S of **words or** symbols into **the sequence** P, and a search algorithm for **determining** the degree of **matching** between **words** or **word sequences in** a **suffix tree** representation of respectively the text T and a query Q. The algorithm searches the data...

...retrieval generates a word-spaced sparse suffix tree for storing suffixes of words in a **text T** as **word sequence** information,and **a**

word size-dependent edit **distance** metric for **word sequences** S, P **and including word** -weighted cost functions for edit operations, and **determines** matches between **word sequences** SR or **retrieved information** R and **word sequences** PQ of a query Q **by calculating** the **edit distance** for all matches. Use in an approximate search **engine**
.

...

...non-evenly spaced sparse suffix tree for storing suffixes of words and/or symbols, or **sequences** thereof, in a **text** T, a metric M including combined **edit distance** metrics for **an** approximate degree of matching respectively between **words and** /or symbols, or between **sequences** thereof, in the **text** T and a query Q, the **latter** distance metric including **weighting** cost functions for edit operations which transform a **sequence** S of the **text** into a **sequence** P of the **query** Q, and search **algorithms** for **determining the** degree of matching respectively between words and/or symbols, or between sequences thereof, in respectively...

...a data structure for storing a text T, a combined metric M which includes an **edit distance** metric for approximate degree of matching between words and/or symbols or **sequences thereof in** the **text** T and words and/or **symbols** in a **sequence** P, **weighting** cost functions for **edit** operations which **transform** a **sequence** S of **words or** symbols into the sequence P, and a search algorithm for **determining** the degree **of** matching between **words** or **word sequences** in a suffix tree representation **of** respectively the text T and **a** query **Q** . **The** algorithm searches the data structure with the query Q, retrieving information with specified match to...

...retrieval generates a word-spaced sparse suffix tree for storing suffixes of words in a **text** T as **word sequence** information,and a **word** size-dependent **edit distance** metric for **word sequences** S, **P and** including **word** -weighted **cost** functions for **edit operations** , and **determines matches between word sequences** SR or **retrieved** information R and **word sequences** PQ of **a** query Q **by calculating** the **edit distance** for all matches. Use in an **approximate search** engine.

...

...des fonctions de ponderation des mots dans des operations de mise au point de textes **transformant** une **sequence** S de mots ou de symboles en une sequence P; et un algorithme de recherche...

...de concordance entre des mots ou sequences de mots dans une representation presentant respectivement le **texte T** et la question Q. L'algorithme **recherche** la structure de donnees en posant la question Q et recupere l'information correspondant specifiquement...
...mots d'information, recourant a un metrique de mesure du niveau de concordance entre les **sequences** de mots **S** et P, comportant des fonctions de cout ponderees en mots pour les operations de mise
Claims:
...und Sequenzen P hieraus umfasst und wobei die Information R Worter und/oder Symbole und **Sequenzen** hieraus aus dem **Text** T umfasst, wobei das Suchsystem eine Datenstruktur zum Speichern wenigstens eines Teils des **Textes** T und ein **Mass** M umfasst, welches den Grad der Ubereinstimmung zwischen der Anfrage Q und der wiedergewonnenen Information...

...fur einen ungefahren Grad an Ubereinstimmung zwischen Wortern und/oder Symbolen s;q in dem **Text** T bzw. einer Anfrage Q und ein Edit-Abstandsmass Dws(S,P) fur einen ungefahren...

...Gewicht-Kostenfunktion fur Editieroperationen umfasst, welche Sequenzen von Wortern und/oder Symbolen s in dem **Text** T in die **Sequenz** P von Wortern und/oder Symbolen q in der Anfrage Q transformiert, wobei die Gewichtung...

...symbols q and sequences P thereof, and retrieved information R comprising words and/or symbols **and** sequences thereof from **the** text T, wherein the search system comprises a data structure for storing at least ...

...words and/or symbols s and sequences S thereof in the text T, that **the metric** M **comprises a** combination of an edit distance metric D(s,q) for an approximate degree of matching between words and/or symbols **s ; q** in respectively the **text** T and **a query** Q and an edit distance metric Dws(S,P) for **an** approximate degree **of** matchingbetween sequences S of words and/or symbols s in the **text** T and **a** query sequence P of words and/or symbols q in **the query** Q, the latter edit distance metric including weighting cost **functions** for **edit** operations which transform sequences of words **and** /or symbols s **in** the text **T** into the sequence P of words and/or symbols q in the query Q,...

...value proportional to a change in the length of the sequence S upon a transformation **or** dependent on **the** size of the **words** and/or **symbols** s;q in sequences S;P to be matched, that the implemented search algorithms **comprise a** first algorithm for determining the degree of matching between words and/or symbols s;q in the suffix tree representation of respectively the text T and a **query** Q, and a second **algorithm** for determining **the degree** of matching between sequences S;P of words and/or symbols s;q **in** the **suffix** tree representation of respectively the text T and the query Q, said first and/or second algorithms searching the **data structure** with **queries** Q **in** the **form** of either words, symbols, sequences of words or sequences of symbols or **combinations thereof** , such that information R is retrieved on the basis of **query** Q **with** a specified degree of matching between the former and the latter, and that the **search** algorithms optionally also **comprise** a third algorithm for determining exact matching between words and/or symbols s;qin the suffix tree representation of respectively the text T and **the** query Q and/ **or a** fourth algorithm **for** determining exact matching between sequences S;P of words and/or symbols s;q in...

...text T and the query Q, said third and/or fourth algorithms searching the data **structure** with **queries** Q **in** the **form** of either words, **symbols** , sequences of words, or sequences of symbols or combinations thereof, such that information R is...

...Systeme **de** recherche **pour** la **recuperation** d' **informations** , en particulier d' **informations** stockees sous forme de texte, dans lequel un texte T comprend des mots et/ou...

...correspondance entre la demande Q et les informations recuperees R, et dans lequel le systeme **de** recherche **met** en **oeuvre** des **algorithmes** de recherche **pour executer** une recherche, en particulier une recherche de texte seulement sur la base de mots clefs...

...distance d'edition DWS(S,P) donnant un degre approximatif de

correspondance entre des sequences **S** de mots et/ **ou** de symboles s du texte T et une sequence de demande P de mots et...

...comprenant des fonctions de ponderation de cout pour des operations d'edition qui transforment des **sequences** de mots **et** /ou de symboles s du texte Ten des sequences P de mots et/ou de...

...et de la demande Q, lesdits premier et/ou deuxieme algorithmes recherchant dans la structure **de** donnees des demandes **Q** ayant la forme de mots, de symboles, de sequences de mots ou ...ces elements, de telle maniere que les informations R sont recuperees sur la base d' **une** demande Q avec **un** degre predetermine de correspondance entre les informations et la demande; etles algorithmes de recherche...

...pour determiner une correspondance exacte entre des mots et/ou des symboles s,q dans **la** **representation** sous forme d'arbre de suffixes respectivement **du** texte T **et** de **la** demande Q et/ou un quatrieme algorithme pour determiner une correspondance exacte entre des sequences S, P de mots et/ou de symboles s,q dans **la** representation **sous** forme d'arbre de suffixes du texte T et de la demande Q, les troisieme...

...A search system for information retrieval, **particularly** information stored **in** form of text, wherein a text T comprises words and/or symbols s and sequences...

...wherein the query Q comprises words and/or symbols q and sequences thereof P, and **retrieved** information R comprising **words** and/or **symbols** and sequences thereof from the text T, wherein the search system comprises a data structure...

...form of a non-evenly spaced sparse suffix tree ST(T) for storing suffixes of **words** and/ **or** symbols s **and** sequences S thereof in the text T, **that** **the** metric M comprises a combination of an edit distance metric D(s,q) for an approximate degree of matching between words and/or symbols s;q in respectively **the** **text** T and a query Q and an edit distance metric Dws(**S** ,P) for **an** approximate degree of matching between sequences S of words and/or **symbols** s in 1the text T and a query sequence P of words and/ **or** **symbols** q in the query Q, the latter edit distance metric **including** weighting **cost** functions for edit operations which transform **sequences** of **words** and/ **or** **symbols** s **in** **the** text T into the sequence P of words and/or symbols q in the query...
...with a value proportional to a change in the length of the sequence S upon **a** transformation **or** dependent on the size **of** the words **and** /or symbols s;q **in** sequences S;P **to** be matched, that **the** implemented search algorithms comprise a first algorithm for determining the degree of matching between words and/or symbols s;q in the suffix tree representation of respectively the text T **and** a query Q, and a **second** algorithm for determining **the** degree of matching between sequences S;P of words and/or symbols s;q in...

...text T and the query Q, said first and/or second algorithms searching the data **structure** with **queries** Q **in** **the** **form** of either words, symbols, sequences of words or sequences of symbols or combinations thereof, such that information R is retrieved **on** **the** basis of query Q with a specified degree of matching between the former and the latter, and that the search algorithms optionally also comprise a third algorithm for **determining** **exact** matching between words and/or symbols s;q in the suffix tree representation of **respectively** the text **T** and **the** query Q and/ **or** **a** fourth algorithm **for** determining **exact** matching between **sequences** S;P of words and/or symbols s;q in **the** **suffix** tree

representation of respectively the text T and the query Q , said third and/or fourth algorithms searching the data structure with queries Q in the form of either words, symbols , sequences of words, or sequences of symbols or combinations thereof, such that information R is

27/69,K/15 (Item 15 from file: 350)
DIALOG(R)File 350:Derwent WPIX

0007594261 - Drawing available
WPI ACC NO: 1996-211301/199622
XRPX Acc No: N1996-176793
String matching method for matching between query and candidate strings - calculating lower bound estimate of string edit distance **between two strings by equalising lengths of strings by adding padding elements to shorter one, sorting elements and summing substitution costs between corresp. elements**
Patent Assignee: HEWLETT-PACKARD CO (HEWP)
Inventor: HULL R
Patent Family (8 patents, 5 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update	
EP 709801	A1	19960501	EP 1994307977	A	19941028	199622	B
JP 8185482	A	19960716	JP 1995237502	A	19950918	199638	E
CN 1131302	A	19960918	CN 1995115969	A	19951024	199801	E
US 5761538	A	19980602	US 1995499967	A	19950710	199829	E
EP 709801	B1	19991229	EP 1994307977	A	19941028	200005	E
DE 69422406	E	20000203	DE 69422406	A	19941028	200013	E
			EP 1994307977	A	19941028		
JP 3067980	B2	20000724	JP 1995237502	A	19950918	200040	E
CN 1098504	C	20030108	CN 1995115969	A	19951024	200532	E

Priority Applications (no., kind, date): EP 1994307977 A 19941028

Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
EP 709801	A1	EN	14	9	
Regional Designated States,Original: DE FR GB					
JP 8185482	A	JA	11		
EP 709801	B1	EN			
Regional Designated States,Original: DE FR GB					
DE 69422406	E	DE			Application EP 1994307977
					Based on OPI patent EP 709801
JP 3067980	B2	JA	12		Previously issued patent JP 08185482

Alerting Abstract EP A1

The matching method calculates a lower bound estimate of the string **edit distance** (46) between the strings by equalising the lengths of the two strings by adding padding elements to the shorter string. The query string and the candidate string (43) are sorted according to the value of their elements. The sum of the substitution costs of the elements in corresp. positions in the two strings is calculated. The candidate best match is stored.

The lower bound distance for each of the remaining candidate strings is calculated. If it is greater than the current one, then that string is discarded. The current best match string or the current candidate string is retained depending on which one has the lower string **edit distance** from the query string as the current best match candidate string. This string is then stored.

USE - In string matching using lower bound estimate of string **edit distance** .

ADVANTAGE - Provides improved method of string matching reducing amount of computation required to perform string matching.

Title Terms/Index Terms/Additional Words: STRING; MATCH; METHOD; QUERY; CANDIDATE; CALCULATE; LOWER; BOUND; ESTIMATE; EDIT; DISTANCE; TWO; EQUAL; LENGTH; ADD; PAD; ELEMENT; SHORT; ONE; SORT; SUM; SUBSTITUTE; COST; CORRESPOND

Class Codes
International Classification (Main): G06K-009/62, G06K-009/68, G06K-009/72
(Additional/Secondary): G06F-017/30, G06F-019/00, G06K-009/00, G06T-007/00
US Classification, Issued: 395899000, 364400000, 382186000, 382187000, 382229000

File Segment: EPI;
DWPI Class: T01; T04
Manual Codes (EPI/S-X): T01-J05B3; T04-D07E

...**calculating lower bound estimate of string** edit distance **between two strings by equalising lengths of strings by adding padding elements to shorter one**...

Alerting Abstract ...The matching method calculates a lower bound estimate of the string **edit distance** (46) between the strings by equalising the lengths of the two strings by adding padding...

...or the current candidate string is retained depending on which one has the lower string **edit distance** from the query string as the current best match candidate string. This string is then...

...USE - In string matching using lower bound estimate of string **edit distance** .

Original Publication Data by Authority

Original Abstracts:
...a query string against a plurality of candidate strings replaces a highly computationally intensive string **edit distance** calculation with **a less** computationally intensive lower bound estimate. The lower bound estimate of the string **edit distance** between the **two strings** is calculated by equalising the lengths of the two strings by adding padding elements to...

...a query string against a plurality of candidate strings replaces a highly computationally intensive string **edit distance** calculation with a less computationally intensive lower **bound estimate** . The lower bound estimate of the string **edit distance** between the two strings is calculated by **equalising the** lengths of the two strings by adding padding elements to the shorter one. The elements...
Claims:
...against a candidate string, the method comprising calculating a lower bound estimate of the string **edit distance between the strings** and **characterised** in **that** calculating the lower bound estimate comprises:</br> - equalising the lengths of the two strings by adding...

...one or more candidate strings, wherein this matching is determined in accordance with a string **edit distance** , wherein the string **edit distance** is a minimum sum of costs of edit operations required to transform one **string** into another; **characterised** in that the step **of** matching the query string against **the one** or more candidate **strings comprises** the step of calculating the lower bound estimate of the string

edit distance between the query **string** and at least one of the candidate strings, wherein the step of calculating the lower bound estimate of the string **edit distance** between the query string and a **candidate string** comprises:</br> - equalising the lengths of the two strings by adding padding elements to the shorter string (63);</br> - sorting the two strings according to the **value of** their internal elements (64);</br> - calculating the sum of the substitution costs of the elements in...

...65), the sum of the substitution costs being the lower bound estimate of the string **edit distance** .

...

...A method for calculating a lower bound estimate of string edit distance between query string and a candidate string, the method comprising:equalising lengths of the strings...

...the query string and the candidate string according to their element values;calculating a sum **of substitution** costs of the elements in corresponding positions in the sorted strings, the sum of the substitution costs being the lower bound estimate of string **edit distance**.>

27/69,K/12 (Item 12 from file: 350)
DIALOG(R)File 350:Derwent WPIX

0010314142
WPI ACC NO: 2000-628288/200060
XRPX Acc No: N2000-465500
Spelling correction method for computer documents by comparing edited words generated with words stored in dictionary and substituting misspelt with unique candidate word
Patent Assignee: JUSTSYSTEM PITTSBURGH RES CENT INC (JUST-N)
Inventor: KANTROWITZ M
Patent Family (2 patents, 87 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update	
WO 2000057291	A1	20000928	WO 2000US260	A	20000106	200060	B
AU 200024922	A	20001009	AU 200024922	A	20000106	200103	E

Priority Applications (no., kind, date): US 1999275701 A 19990324

Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
WO 2000057291	A1	EN	57	0	

National Designated States,Original: AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CR CU CZ DE DK DM EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG UZ VN YU ZA ZW
Regional Designated States,Original: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SL SZ TZ UG ZW

AU 200024922	A	EN		Based on OPI patent	WO 2000057291

Alerting Abstract WO A1
NOVELTY - The minimum **edit distance** is calculated using a restricted set of edit operations which correct common errors comprising insertion, deletion, transposition and/or substitution after comparison with a dictionary of valid words. A unique candidate word for misspelled word is substituted after comparing edited word generated in the preceding step.
USE - For computer documents.
ADVANTAGE - It reduces the time required for correcting spelling error.

Title Terms/Index Terms/Additional Words: SPELLING; CORRECT; METHOD; COMPUTER; DOCUMENT; COMPARE; EDIT; WORD; GENERATE; STORAGE; DICTIONARY; SUBSTITUTE; MISSPELT; UNIQUE; CANDIDATE

Class Codes
International Classification (Main): G06F-015/00
(Additional/Secondary): G06F-017/00, G06F-017/21, G06F-017/24, G06F-017/27, G07F-011/08

File Segment: EPI;
DWPI Class: T01
Manual Codes (EPI/S-X): T01-J; T01-J11A1; T01-S01B

Original Titles:
SPELLING CORRECTION METHOD USING IMPROVED MINIMUM **EDIT DISTANCE** ALGORITHM...
...NOVELTY - The minimum **edit distance** is calculated using a restricted set of edit operations which correct common errors comprising insertion...

Original Publication Data by Authority

Original Abstracts:

...method of spelling correction comprises the steps of: a) storing a dictionary of valid words, **b**) for each **input string** to be **checked** comparing the input **string to words** in **the** stored **dictionary** to **identify** input **strings** not **in** the **dictionary** , c) for each input string not found in the preceding step, generating test words by...

...operations which correct the most common errors comprising insertion, deletion, transposition and/or substitution, d) **comparing** the edited input **string** generated in the **preceding** step with **words** stored in the **dictionary** and e) generating a candidate word or candidate list of the words...

10/69,K/3 (Item 3 from file: 350)
DIALOG(R)File 350:Derwent WPIX

0007769613 - Drawing available
WPI ACC NO: 1996-395197/199640
XRPX Acc No: N1996-333044
Computer implemented method for identifying patterns in transaction sequences - involves identifying which large sequences are not subsets of other large sequences and returning maximal large sequences to indicate recurring purchasing patterns over time
Patent Assignee: IBM CORP (IBMC); INT BUSINESS MACHINES CORP (IBMC)
Inventor: AGRAWAL R; SRIKANT R
Patent Family (6 patents, 4 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update	
EP 730240	A2	19960904	EP 1996301405	A	19960301	199640	B
JP 8263346	A	19961011	JP 199638077	A	19960226	199651	E
US 5819266	A	19981006	US 1995398640	A	19950303	199847	E
EP 730240	B1	20020612	EP 1996301405	A	19960301	200239	E
DE 69621670	E	20020718	DE 69621670	A	19960301	200255	E
			EP 1996301405	A	19960301		
JP 3373716	B2	20030204	JP 199638077	A	19960226	200317	E

Priority Applications (no., kind, date): EP 1996301405 A 19960301; US 1995398640 A 19950303

Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
EP 730240	A2	EN	18	9	
Regional Designated States,Original: DE FR GB					
JP 8263346	A	JA	19		
EP 730240	B1	EN			
Regional Designated States,Original: DE FR GB					
DE 69621670	E	DE			Application EP 1996301405
					Based on OPI patent EP 730240
JP 3373716	B2	JA	16		Previously issued patent JP 08263346

Alerting Abstract EP A2
The method involves entering an item set into a large set of item sets when the number of items is present in the database exceeds a predefined minimum support value. A forward set of large sequences is defined and sequences are concatenated in accordance with a predetermined concatenation regime to generate a next set of candidate large sequences.
Each sequence in the next set of candidate large sequences is compared to the sequences in a transformed set of sequences. A candidate large sequence is entered into a next forward set. The set of maximal large sequences is output for identifying particular transaction sequences over time.
ADVANTAGE - Allows for quick mining of large databases which is easy to use and cost-effective.

Title Terms/Index Terms/Additional Words: COMPUTER; IMPLEMENT; METHOD; IDENTIFY; PATTERN; TRANSACTION; SEQUENCE; SUBSET; RETURN; MAXIMUM; INDICATE; RECURRENCE; PURCHASE; TIME

Class Codes
International Classification (Main): G06F-012/00, G06F-017/30, G06F-017/60, G06F-009/44
(Additional/Secondary): G06F-019/00

US Classification, Issued: 707006000

File Segment: EPI;
DWPI Class: T01
Manual Codes (EPI/S-X): T01-J05A; T01-J05B4

Original Publication Data by Authority

Claims:
...large sequences when the number of times the candidate large sequence is present in the **transformed set** of sequences is **greater** than the minimum support value; and</br> (**f**) outputting the set of maximal large sequences for identifying particular transaction sequences over time...

...38) an itemset into a set of large itemsets when the number of times the **itemset** is present in the database **exceeds** a predefined value; **characterised** by:generating (40) a transformed **set** of sequences by discarding a transaction when the transaction does not include an itemset in...

...does not include an itemset in the set of large itemsets; andgenerating (42) a **set** of **large sequences** which appear in the database a number of times exceeding the predefined value, including the ...large sequences and concatenating sequences in the forward set of large sequences to generate a **next set** of candidate **large** sequences;(**d**) comparing each sequence in the next set of candidate large sequences to the sequences in t ...

15/69,K/1 (Item 1 from file: 350)
DIALOG(R)File 350:Derwent WPIX

0012257261 - Drawing available
WPI ACC NO: 2002-197363/200226
XRPX Acc No: N2002-149932
Molecular sequence aligning method for microbiological applications, involves computing gap effect in individual or prealigned multiple sequences, and symbol jumping effects in prealigned sequence
Patent Assignee: DEUT KREBSFORSCHUNGSZENTRUM (DEKR-N)
Inventor: SPANG R; STOYE J
Patent Family (1 patents, 25 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
EP 1152349	A1	20011107	EP 2000109664	A	20000506	200226 B

Priority Applications (no., kind, date): EP 2000109664 A 20000506

Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
EP 1152349	A1	EN	29	10	

Regional Designated States,Original: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI

Alerting Abstract EP A1

NOVELTY - The matching or coincidence between symbols of individual molecular sequence, and prealigned multiple molecular sequences are determined at evaluation position. The effect of introducing a gap as a new position in any one sequence, and the effect of symbol jumping from one prealigned sequence to the other are computed to determined the optimum individual alignment.

USE - For aligning of sequences e.g. nucleotide and peptide sequences in microbiology.

ADVANTAGE - The computation of the gap introducing effects, the alignment of the molecules are simplified. The time complexity associated with the calculation of jumping effects is reduced, by calculating the jump effects from one sequence to the other, which is the same for all sequences. Hence, the need for calculating the best jump individually for each sequence is eliminated.

DESCRIPTION OF DRAWINGS - The figure shows the graphical representation of an **edit matrix** .

Title Terms/Index Terms/Additional Words: MOLECULAR; SEQUENCE; ALIGN; METHOD; MICROBIOLOGICAL; APPLY; COMPUTATION; GAP; EFFECT; INDIVIDUAL; MULTIPLE; SYMBOL; JUMP

Class Codes
International Classification (Main): G06F-017/30

File Segment: EPI;
DWPI Class: T01
Manual Codes (EPI/S-X): T01-J04C; T01-J06A; T01-S01C

Alerting Abstract ...DESCRIPTION OF DRAWINGS - The figure shows the graphical representation of an **edit matrix** .

Original Publication Data by Authority

Original Abstracts:
...candidate sequence and a number of prealigned seed sequences; wherein

the sequences are defined by **a** consecutive **string** of **characters** , each **of** which **belongs** to a **group** of symbols including a gap symbol; with the following step: -Finding an optimal alignment between...

...the candidate sequence the best fit with a corresponding position of one of the seed **sequences** ; **characterised** in that **for determining** the optimal individual **alignment** at each position under evaluation, the following parameters are used: -the match between the symbol...

Claims:

...and a number of prealigned seed sequences; wherein the sequences are defined by a consecutive **string** of **characters** , each **of** which belongs to a **group** of symbols including a gap symbol; with the following step: finding an optimal alignment between...

...the candidate sequence the best fit with a corresponding position of one of the seed **sequences** ; **characterised** in that **for determining** the optimal **individual** alignment at each position **under evaluation** , the following parameters are used: the match between the symbol of the candidate sequence and...

27/69,K/10 (Item 10 from file: 350)
DIALOG(R)File 350:Derwent WPIX

0010953596 - Drawing available
WPI ACC NO: 2001-576697/200165
Method for calculating phonetic distance among different phonetic spelling of adopted word
Patent Assignee: KOREA ADV INST SCI & TECHNOLOGY (KOAD)
Inventor: CHOI G S; CHOI K; KANG B; KANG B J
Patent Family (3 patents, 2 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update	
KR 2001035679	A	20010507	KR 199942372	A	19991001	200165	B
KR 318762	B	20020104	KR 199942372	A	19991001	200253	E
US 6581034	B1	20030617	US 2000483860	A	20000117	200341	E

Priority Applications (no., kind, date): KR 199942372 A 19991001

Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
KR 2001035679	A	KO	1	10	
KR 318762	B	KO			Previously issued patent KR 2001035679

Alerting Abstract KR A
NOVELTY - A phonetic distance calculation method is provided to perform a phonetic distance among different spellings on an adopted word so that it can enhance an efficiency of an information retrieval system.
DESCRIPTION - The method comprises steps of a system manager defining consonant and vowel change patterns which can occur among the different phonetic spellings on an adopted word(S1), the system generating the different phonetic spellings based on the defined consonant and vowel change patterns and granting a penalty on each different phonetic spelling(S2), the system calculating a minimum phonetic distance between the two different phonetic spellings by applying a minimum **edition distance** calculation method(S3), and the system determining the different phonetic spelling with the shortest phonetic distance(S4). The consonant and vowel change patterns is classified into a transposition of the consonants and vowels, an insertion/deletion of the consonants and vowels, and an extension/compression of the consonants and vowels.

Title Terms/Index Terms/Additional Words: METHOD; CALCULATE; PHONETIC; DISTANCE; SPELLING; ADOPT; WORD

Class Codes
International Classification (Main): G06F-017/20, G10L-015/10
(Additional/Secondary): G06F-017/21
US Classification, Issued: 704238000, 704010000

File Segment: EngPI; EPI;
DWPI Class: T01; P86
Manual Codes (EPI/S-X): T01-J11

Alerting Abstract ...calculating a minimum phonetic distance between the two different phonetic spellings by applying a minimum **edition distance** calculation method(S3), and the system determining the different phonetic spelling with the shortest phonetic...

Original Publication Data by Authority

Original Abstracts:

A phonetic distance calculation method for similarity comparison between phonetic transcriptions **of** foreign words. A system manager **defines** **character** element transformation **patterns** **occurrable** between phonetic **transcriptions** derived from the same foreign language. A system generates new phonetic transcriptions according to the defined **character** element transformation **patterns** **and** assigns a **demerit** mark to each of the generated phonetic transcriptions according to a phonetic distance. A minimum...

...phonetic transcriptions and a given phonetic transcription is calculated on the basis of a minimum **edit** **distance** calculation method. **Any** **one** of the generated phonetic transcriptions with a smallest one of the calculated minimum phonetic distances...

Claims:

...1. A phonetic distance calculation method for similarity comparison between phonetic transcriptions of foreign words, **comprising** the steps of: a) defining **character** element transformation **patterns** occurrable between phonetic transcriptions derived **from** the same **foreign** language, by classifying said patterns into three types, substitution of one **character** element **with** a different one, insertion or deletion of one character element, and expansion of one character...

...into two or contraction of two consecutive character elements into one, wherein said expansion/contraction **patterns** **are** restricted to **character** element **combinations** as following **table** , classifying said three **types** into **consonants** and vowels, and then classifying said consonants into final and initial consonants, **Patterns** **Character** element **combinations** Final/initial + */ E-, -*/ -, -*/ -, -*/ -, vowel -*/ --, -*/ - Vowel/vowel + vowel -/ -, -/ -, -/ -, -, -, -/ -, -/ -, -/ - assigning a demerit mark to each of said patterns, and storing said patterns and said **demerit** marks into **a** database; **b**) taking a phonetic transcription input, and generating new phonetic transcriptions for said input and assigning...

...between each of said new phonetic transcriptions and said input on basis of a minimum **edit** **distance** calculation method, wherein using said demerit marks; and d) considering any one of said new

```
Set     Items   Description
S1      87348   (WORD??? OR CHARACTER? OR TERM? ? OR TEXT? ? OR ALPHABET? -
             OR SPELLING? OR LETTER?)(3N)(STRING? OR SEQUEN? OR PATTERN? OR
              COMBINATION?)
S2       5678   S1(7N)(MONITOR? OR EXAMIN? OR DETECT? OR UNCOVER? OR REVEA-
             L? OR ASSESS? OR EVALUAT? OR INSPECT?)
S3      13414   S1(7N)(DETERMIN? OR COMPAR? OR DISCERN? OR ASCERTAIN? OR A-
             NALY? OR IDENT? OR CHECK? OR VERIF? OR JUDG???)
S4       8226   S1(5N)(SET? ? OR GROUP? OR CLUSTER? OR BUNCH? OR COLLECTIO-
             N? OR AGGREGAT? OR SAMPL? OR DICTIONAR?)
S5     380681   (MAXIMUM? OR MAX OR TOP OR HIGH? OR GREAT? OR PEAK??? OR P-
             ENULTIMAT? OR FULL??? OR MOST?? OR LAST?? OR THRESHOLD? OR TH-
             RESHHOLD? OR BRINK)(5N)(NUMBER? OR AMOUNT? OR COUNT? OR SUM OR
              TOTAL? OR TALLY? ? OR QUANTIT? OR MANY OR KNOWN OR SIZE? ?)
S6      10341   S5(7N)(EDIT??? OR LEVENSHTEIN? OR HAMMING? OR PARE? ? OR P-
             ARING OR INSERT? OR DELET? OR TRANSFORM? OR CHANG???)
S7      97302   (STOP??? OR STOPPING OR CEASE?? OR CEASING OR END??? OR HA-
             LT??? OR CESSAT?)(3N)(COMPUT? OR DETERMIN? OR TOTAL? OR TABUL-
             AT? OR CALCULAT? OR PROCESS? OR FIGUR? OR FORMULAT?)
S8       2755   S2:S3 AND S4
S9        176   S8 AND S5
S10         5   S9 AND S6
S11        28   S9 AND (MATRIX? OR GRID? OR TABLE? OR MATRIC?)
S12         0   S11 AND (EDIT??? OR LEVENSHTEIN?)
S13        98   S8 AND (EDIT??? OR LEVENSHTEIN?)
S14        11   S13 AND (MATRIX? OR GRID? OR TABLE? OR MATRIC?)
S15        11   S14 NOT S10
File 350:Derwent WPIX 1963-2007/UD=200762
         (c) 2007 The Thomson Corporation
File 347:JAPIO Dec 1976-2007/Jun(Updated 070926)
         (c) 2007 JPO & JAPIO
```